Ribosomal Protein Gene Clusters and Ribosome Evolution

Indrani Dasgupta, Yamei Liu , Jiachen Wang, Hung-Chung Huang, Uma Nagaswamy, and **George E. Fox**
Department of Biology & Biochemistry, University Houston, Houston, TX 77204-5001

The emergence of the translation machinery likely was the event that catalyzed the transition from the RNA World to the protein world that now characterizes life as we know it. The ribosomal machinery is complex with multiple components that are probably not of equal age. Instead, they have likely been recruited to and from the machinery over evolutionary time. We are using bioinformatics approaches to unravel as much of this history as possible. Our initial studies suggested that the order of assembly of the 50S ribosomal subunit likely recapitulates its history with the earliest proteins that are incorporated likely being older than those that follow. A recent examination of the revised 30S assembly map illustrates how the RNA and associated protein components may have increased in complexity together.

The conclusion that the core ribosomal proteins (r-proteins) in the assembly process are likely to be older is supported by the fact that they are typically universal in their distribution while the proteins added late in assembly are frequently not universally distributed. In addition to their conservation, the large subunit proteins hypothesized to be older belong to one of two large r-protein gene clusters, i.e. the spc or S10 operon. These two clusters are widely distributed among all Bacteria and Archaea. In *E. coli* they are known to be operons where they are subject to autogenous control which may in fact be the earliest mechanism for regulating transcription. In this process, depending on the availability of rRNA, one of the proteins in the operon binds directly to either the rRNA or a region of the mRNA that mimics the rRNA binding site. If these regulatory systems are in fact highly conserved, they may also speak to a time of origin relationship between the protein cluster and the relevant regions of the rRNA. In general, the relative age of various ribosomal and other cellular components may be reflected in when and to what extent these regulatory relationships have emerged and persisted.

In order to explore this further, we examined 100 phylogenetically representative genomes from Bacteria and Archaea to determine the extent to which clusters of ribosomal proteins and other ribosomal components have been conserved over evolutionary time. The immediate neighbors of every widely distributed ribosomal component were tabulated and approximately 20 recurring clusters of proteins were identified. Many of the clusters have alternative components but typically there is a common core. The relative conservation of each core cluster was examined and seven were extremely conserved. One of these contains rpoA indicating a very early coupling between transcription and translation. DNA replication machinery components do not occur in any except the most variable clusters. With regard to the 30S subunit, all the proteins in the S7 assembly path belong to the most conserved clusters. This is consistent with the notion that the decoding site is at the historical core of the 30S subunit. Ongoing work is focused on determining the extent to which potential RNA regulatory structures are preserved in the conserved clusters.